

AD 693 116

DIGITAL COMPUTER SIMULATION: ESTIMATING SAMPLE
SIZE

George S. Fishman

Rand Corporation
Santa Monica, California

August 1969

AD 693116

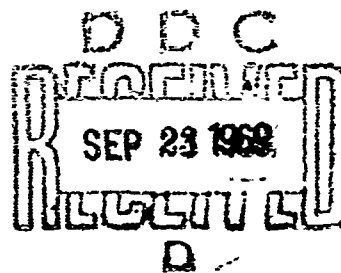
MEMORANDUM

FM-5866-PR

AUGUST 1969

DIGITAL COMPUTER SIMULATION: ESTIMATING SAMPLE SIZE

George S. Fishman



PREPARED FOR:

UNITED STATES AIR FORCE PROJECT RAND

The **RAND** *Corporation*
SANTA MONICA • CALIFORNIA

THIS DOCUMENT HAS BEEN APPROVED FOR PUBLIC RELEASE AND SALE; ITS DISTRIBUTION IS UNLIMITED.

61

MEMORANDUM

RM-5866-PR

AUGUST 1969

DIGITAL COMPUTER SIMULATION:
ESTIMATING SAMPLE SIZE

George S. Fishman

This research is supported by the United States Air Force under Project RAND—Contract No. F44620-67-C-0015—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of the United States Air Force.

DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

The RAND Corporation

1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

This study is presented as a competent treatment of the subject, worthy of publication. The Rand Corporation vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the authors.

Published by The RAND Corporation

PREFACE

This Memorandum continues RAND's research into the statistical analysis of computer simulation experiments. The overall purpose is to find methods for efficiently extracting useful information from time series generated by these experiments. This particular study describes a technique for estimating the required sample size in a simulation experiment, and provides flow charts and computer programs for incorporating the proposed technique directly into a computer simulation program. Emphasis is on relieving the investigator of the need to interact with the ongoing simulation to determine when the desired statistical precision has been obtained.

Preceding work on this subject is described in G. S. Fishman and P. J. Kiviat, Spectral Analysis of Time Series Generated by Simulation Models, The RAND Corporation, RM-4393-PR, February 1965; G. S. Fishman, Problems in the Statistical Analysis of Simulation Experiments: The Comparison of Means and the Length of Sample Records, The RAND Corporation, RM-4880-PR, February 1966; and G. S. Fishman, Digital Computer Simulation: The Allocation of Computer Time in Comparing Simulation Experiments, The RAND Corporation, RM-5288-1-PR, October 1967.

SUMMARY

A method will be described for estimating and collecting the sample size needed to evaluate the mean of a process (with a specified level of statistical accuracy) in a simulation experiment. A procedure is also described for incorporating the determination and collection of the sample size into a computer library routine that can be called by the ongoing simulation program.

We present the underlying probability model that enables us to denote the variance of the sample mean as a function of the autoregressive representation of the process under study. And we describe the estimation and testing of the parameters of the autoregressive representation in a way that can easily be "built into" a computer program.

Several reliability criteria are discussed for use in determining sample size. Since these criteria assume that the variance of the sample mean is known, an adjustment is necessary to account for the substitution of an estimate for this variance. It is suggested that Student's distribution be used as the sampling distribution, with "equivalent degrees of freedom" determined by analogy with a sequence of independent observations.

A bias adjustment is described that can be applied to the beginning of the collected data to reduce the influence of initial conditions on events in the experiment. Four examples are presented using these techniques, and comparisons are made with known theoretical solutions. Finally, we present the minimum variance unbiased estimator of the sample mean, which turns out to be a function of the

autoregressive coefficients. Before these results can be used in practice, more will have to be known about their sampling properties.

In conclusion, it is noted that the use of the procedures described here relieves the user of the task of continually interacting with the simulation experiment to determine whether his results are within an acceptable range.

CONTENTS

PREFACE	iii
SUMMARY	v
Section	
1. INTRODUCTION	1
2. THE MODEL	6
3. STATISTICAL INFERENCE	21
4. RELIABILITY CRITERIA	25
5. IMPROVED CONFIDENCE INTERVALS	30
6. BIAS ADJUSTMENT	32
7. EXAMPLES	35
8. THE MINIMUM VARIANCE ESTIMATOR	47
9. CONCLUSIONS	54
REFERENCES	55

1. INTRODUCTION

In many simulation experiments, observations collected on the process of interest are positively correlated. This means that an observation that exceeds (falls below) the population mean of the process tends to be followed by another observation that exceeds (falls below) the mean. In a more limited set of simulation experiments, successive observations are negatively correlated so that an observation exceeding (falling below) the mean tends to follow one falling below (exceeding) it.

Estimating a population mean from sample data is a common objective of the statistical analysis of a simulation experiment and, moreover, an estimate of the variance of the sample mean is helpful in assessing how representative the sample mean is of the population mean. For uncorrelated observations, the sample population variance divided by the sample size provides a convenient estimate of the variance of the sample mean. For correlated data, the variance of the sample mean is a function of the correlation between observations, a fact that causes considerable difficulty in estimating this variance.

Estimators that take account of the correlation to varying extents have been suggested in the literature [4,6], but all require a degree of subjective judgment regarding their adequacy. Ideally, one wants an algorithm that can be "built into" a computer simulation program and can objectively estimate the sample size needed to obtain a specified confidence interval for a population mean. Such a procedure would relieve an investigator of the burden of estimating the variance of the sample mean from a data sample obtained from a trial run, estimating

the sample size necessary for the specified confidence interval, and then collecting that many more observations in a successive simulation run. An ideal program would accomplish these tasks without taking the simulation off the computer. This Memorandum describes an algorithm for doing this.

Estimating the population mean and estimating the variance of this resulting estimate are problems in statistical inference that require an underlying probability model. Some models are more convenient than others, and it is natural to assume a model that yields desirable statistical properties. For example, we may assume that observations are independent and identically distributed if we believe that the outcome of any trial is not influenced by the outcomes of other trials and also if the ordering of the trials does not affect their outcomes.

If we suspect that the observations are statistically dependent but that the dependence is strictly nonlinear, then we may assume that the observations are uncorrelated. This model is less restrictive than that for the independent case, for it implies that the observations share a common mean and a common variance; the covariance between any two observations is zero, but no specification is made regarding the nonlinear statistical relationship among observations.

The assumptions of independent or uncorrelated observations apply in many statistical analyses; in simulation experiments, however, observations are often *autocorrelated*. This means, statistically, that an observation is linearly dependent on preceding observations. Since failure to acknowledge this autocorrelation can seriously impair

the veracity of conclusions based on statistical results, a probability model is needed to account for autocorrelation.

The commonly employed estimator of the population mean for autocorrelated observations has the same algebraic form as that for independent observations. In the independent case, it is statistically unbiased and has minimum variance; in the autocorrelated case, these properties hold only for large samples. Nevertheless, the estimator has much intuitive appeal and can be computed easily, two properties that account for its common use.

The variance of the sample mean is a function of the autocorrelation between observations and, therefore, specification regarding the covariance structure is necessary to make the estimation of this variance a tractable statistical problem. The probability model of a covariance stationary sequence provides a convenient framework within which this problem and many others can be solved. It is this model that we describe in Sec. 2.

In [4], the general properties of a covariance stationary sequence enable us to derive a useful estimator of the variance of the sample mean; but, unfortunately, that estimator is difficult to build into a simulation program. By adding several mild restrictions to the model, we may represent a present observation as a linear combination of past observations plus a random residual uncorrelated with past observations. This scheme is called an *autoregressive representation* of the sequence, and the weights in the linear combination are called the *autoregressive coefficients*. For large samples, it is shown that a knowledge of the autoregressive coefficients and the variance of

the uncorrelated residuals enables us to approximate the variance of the sample mean closely.

While the autoregressive coefficients and the residual variance are unknown, we can estimate them as described in Sec. 3. We then use these to estimate the variance of the sample mean. This approach is desirable because the estimation and testing of the coefficients, the computation of the sample residual variance, and the estimation of the variance of the sample mean can all be accomplished directly in a simulation program and require no user intervention.

Section 4 discusses several criteria for determining sample size. In some experiments we may want the confidence interval to have some fixed absolute width around the mean. In others we may require the width to be a fixed percentage of the mean. To account for the use of an estimate for the variance of the sample mean, Sec. 5 introduces the t distribution with appropriate adjustments for its use with autocorrelated data. Initial conditions often influence the behavior of the process under study; and it is desirable, whenever possible, to reduce the extent of this influence. Section 6 covers this topic.

In Sec. 7, several examples with known solutions are presented to illustrate how well the techniques work statistically. The examples include zero-, first-, and second-order autoregressive schemes and a single-server queueing problem with independent and exponentially distributed interarrival times and service times. These examples are simulated and the estimated sample sizes compared with known theoretical results presented in [5].

Earlier we remarked that the conventional estimator of the mean of an autocorrelated sequence is unbiased and minimum variance only

for large samples. It is therefore instructive to study what can be done to derive an improved estimator of the mean for a moderate sample size and how feasible and worthwhile it is to use the improved estimator. Section 8 presents the minimum variance estimator and compares it with the conventional mean estimator.

We conclude that the algorithms suggested here can contribute significantly to solving the problem of determining sample size in a simulation experiment. Because they can be used while minimally involving the simulation experiment itself, they are worthy of consideration, especially since they can be easily modified to meet individual user preferences. Also, the suggestions regarding reliability criteria, unbiasedness, minimum variance and variance reduction techniques provide users with information that enables them to draw useful inference about the process being studied.

2. THE MODEL

In many simulation experiments, the process of interest appears as a sequence of events in which the index that orders the events may play a role in defining the relationship among events. The index may be time; for example, the number of units waiting for service at time t , where t assumes T different values. The index may simply denote order; for example, the waiting time for the t^{th} unit to receive service, where t assumes the values $1, \dots, N$.

When the set of ordered events is subject to random variation, it is called a stochastic sequence. Let X_t be the value assumed by the t^{th} event in which the index t runs over the integers. Then we denote the stochastic sequence by $\{X_t; t = 0, \pm 1, \dots, \pm \infty\}$ or, more concisely, by X . We could index the events on a finite set of non-negative integers, but the above definition of X offers several expositional conveniences without impairing its applicability in the present context.

Let the sequence X have mean

$$(2.1) \quad \mu_t = E(X_t)$$

and autocovariance function

$$(2.2) \quad R_{s,t} = E[(X_s - \mu_s)(X_t - \mu_t)] .$$

If μ_t is finite and independent of the ordering index t , and $R_{s,t}$ is finite and a function only of the number of intervening events $s-t$, then we may write the mean and autocovariance function as

$$(2.3) \quad \mu = E(X_t) ,$$

$$R_{s-t} = R_{t-s} = E[(X_s - \mu)(X_t - \mu)] ,$$

respectively. A sequence satisfying (2.3) is called covariance, mean-square, weakly or wide-sense stationary.*

Suppose that

$$(2.4) \quad R_{s-t} = \begin{cases} \sigma^2 & s = t \\ 0 & s \neq t \end{cases} .$$

Then X is a sequence of uncorrelated events, and conventional methods of statistical inference apply when estimating μ . In general, (2.4) does not hold and more sophisticated inferential methods are needed.

For a sample of N observations, we compute the conventional sample mean as

$$(2.5a) \quad \bar{X}_N = (1/N) \sum_{t=1}^N X_t ,$$

* See [1] for a more complete description of covariance stationary sequences.

with variance

$$\begin{aligned} \text{Var}(\bar{X}_N) &= (1/N^2) \sum_{s,t=1}^{N-1} R_{s-t} \\ (2.5b) \quad &= (1/N) \sum_{s=1-N}^{N-1} (1 - |s|/N) R_s . \end{aligned}$$

We also require that

$$(2.6) \quad \lim_{s \rightarrow \infty} R_s = 0 .$$

This restriction is reasonable, for we would expect the covariance between events in the series to vanish as the number of intervening events increases. Then one may show that

$$(2.7a) \quad \lim_{N \rightarrow \infty} \sum_{s=1-N}^{N-1} R_s = m < \infty ,$$

$$(2.7b) \quad \lim_{N \rightarrow \infty} \sum_{s=1-N}^{N-1} (|s|/N) R_s = 0 ,$$

so that for large N

$$(2.8) \quad \text{Var}(\bar{X}_N) \sim V_N = m/N .$$

Condition (2.6) requires the absence of any regularly periodic components in X . If any were present, then the autocovariance function R would contain undamped cosine terms that would violate (2.6) and prevent the convergence of (2.7b). The following example demonstrates the truth of this assertion. Let X_t be defined by

$$(2.9) \quad X_t = a \sin bt + y_t ,$$

where a is a random variable with zero mean and unit variance, b is a constant, and y is a covariance stationary sequence with mean v and autocovariance function P . Moreover,

$$(2.10) \quad \lim_{S \rightarrow \infty} P_S = 0.$$

Then X has the autocovariance function

$$(2.11) \quad R_S = \cos bs + P_S$$

and, for a sample of N observations,

$$(2.12) \quad \text{Var}(\bar{X}_N) = (1/N) \sum_{s=1-N}^{N-1} (1 - |s|/N) (\cos bs + P_S) .$$

Now

$$\sum_{s=1-N}^{N-1} \cos bs = \sin [b(2N - 1)/2] / \sin (b/2) ,$$

$$\sum_{s=1-N}^{N-1} |s| \cos bs = N \sin[b(2N-1)/2] / \sin(b/2) - [1 - \cos(Nb)] / [2 \sin^2(b/2)],$$

so that

$$\sum_{s=1-N}^{N-1} (1 - |s|/N) \cos bs = [1 - \cos(bN)] / [2N \sin^2(b/2)].$$

Then

$$(2.13) \quad \lim_{N \rightarrow \infty} \text{Var}(\bar{X}_N) = \sum_{s=-\infty}^{\infty} P_s \neq \sum_{s=-\infty}^{\infty} R_s,$$

so that (2.8) does not hold in this case. This result suggests that any regularly periodic components in the process X be removed prior to estimating the mean μ .

The reader may wonder why (2.8) is of such great significance, since m is an infinite sum of autocovariances. The answer is that m can be computed from alternative formulae wherein the individual R_s 's need not be known. As a result our emphasis is on the quantity m and its estimation. The autoregressive representation, which is to be introduced shortly, provides alternative formulae for computing m .

Condition (2.6) implies that the correlation between two events X_s and X_t goes to zero as the interval $|s-t|$ becomes large. If we impose the added, but mild, restriction that there exists a finite

interval r such that two events, X_s and X_t , are statistically independent if $|s - t| > r$, and that $E(|X_t|^3) < \infty$, then it can be shown that the limiting distribution of the quantity $N^{\frac{1}{2}}(\bar{X}_N - \mu)$ is normal with mean zero and variance σ^2 [7, pp. 215-219]. Hereafter, we assume that N is sufficiently large for us to use this limiting result.

One of the desirable features of a covariance stationary sequence is its connection with a sequence of uncorrelated, identically distributed random variables. Using the Wold decomposition theorem [16], one may write for X_t satisfying (2.3) and (2.6)[†]

$$(2.14) \quad X_t = \mu + \sum_{s=0}^{\infty} a_s Y_{t-s},$$

where $\{a_s; s=0, \pm 1, \pm 2, \dots, \pm \infty\}$ is a sequence of real constants with

$$(2.15) \quad \sum_{s=0}^{\infty} a_s^2 < \infty;$$

and $\{Y_t; t = 0, \pm 1, \pm 2, \dots, \pm \infty\}$ is a sequence of uncorrelated, identically distributed random variables with mean zero and variance σ^2 . This is an appealing form, for we can now write the autocovariance of X as

$$(2.16) \quad R_s = \sigma^2 \sum_{t=0}^{\infty} a_t a_{s+t},$$

[†]For a concise description of the Wold decomposition theorem, see [2, pp. 286-288].

so that

$$(2.17) \quad \sigma^2 = \sum_{s=0}^{\infty} R_s = \sigma^2 \left(\sum_{s=0}^{\infty} a_s \right)^2 .$$

This result is interesting but, since the sequence Y is seldom observed, it is difficult to infer the a_s 's statistically. We may nevertheless benefit from (2.14). Taking z transforms leads to

$$(2.18) \quad \left\{ \begin{aligned} X'(z) &= \sum_{s=0}^{\infty} (X_s - \mu) z^s = \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} a_t Y_{s-t} z^s \\ &= A(z)Y(z) \\ A(z) &= \sum_{s=0}^{\infty} a_s z^s \\ Y(z) &= \sum_{s=0}^{\infty} Y_s z^s . \end{aligned} \right.$$

Note that

$$(2.19) \quad A(1) = \sum_{s=0}^{\infty} a_s ,$$

$$\sigma^2 = \sigma^2 A^2(1) .$$

Under fairly mild conditions on $[A(z)]^{-1}$ we may write[†]

$$(2.20) \quad \left\{ \begin{array}{l} B(z)X'(z) = Y(z) \\ B(z) = 1/A(z) \\ B(z) = \sum_{s=0}^{\infty} b_s z^s. \end{array} \right.$$

On taking the inverse transform we have

$$(2.21) \quad \sum_{s=0}^{\infty} b_s X'_{t-s} = Y_t,$$

which is called the *autoregressive representation* of X . Notice that

$$(2.22) \quad m = \sigma^2/B^2(1) = \sigma^2 / \left(\sum_{s=0}^{\infty} b_s \right)^2.$$

If we can estimate σ^2 and the b_s 's, then we can estimate m . In its present form, (2.21) does not enable us to estimate these quantities by conventional methods. We have already assumed that the zeros of $A(z) = 0$ do not lie on the unit circle. If we also assume that the zeros of $B(z) = 0$ lie outside the unit circle, then one may write (2.21) as

[†]See Whittle [15, pp. 26-27].

$$b_s = 0 \quad s < 0 ,$$

(2.23)

$$\sum_{s=0}^{\infty} b_s X'_{t-s} = Y_t .$$

Moreover, the *moving average* (2.14) is expressible as

(2.24)

$$X'_t = \sum_{s=0}^{\infty} a_s Y_{t-s} .$$

Expression (2.24) has an intuitive appeal, for it implies that X'_t is a linear combination of uncorrelated, identically distributed *present* and *past* events.

The number of coefficients in the b sequence remains to be considered. Suppose that

$$b_s = 0 \quad s > p > 0 ,$$

so that

(2.25)

$$\sum_{s=0}^p b_s X'_{t-s} = Y_t .$$

One would normally expect that after some lag p , the contribution to the behavior of X'_t made by variables X'_{t-p-1} , X'_{t-p-2} , ... would be negligible, so that (2.25) would adequately describe the relationship

between X' and Y . This means that we can find a linear combination of present and past values of X' that form a sequence Y of uncorrelated, identically distributed events with mean zero and variance σ^2 .

Using (2.25) leads to

$$(2.26) \quad m = \sigma^2 / b^2 ,$$

$$b = \sum_{s=0}^p b_s .$$

To estimate σ^2 and the coefficients in the b sequence, we apply the linear least-squares method to a sample of observations on X using the autoregressive representation (2.25). We can subsequently estimate m by substituting estimates of σ^2 and the b_s 's into (2.26). The sample variance of the sample mean is then m/N .

It is instructive to compare m with $N \text{Var}(\bar{X}_N)$ to measure the adequacy of the approximation. Suppose X has the autoregressive representation

$$(2.27) \quad X_t - gX_{t-1} = Y_t \quad |g| < 1 ,$$

so that

$$(2.28) \left\{ \begin{array}{l} R_s = \sigma^2 g |s| / (1-g^2) \\ N \text{ Var}(\bar{X}_N) = 1/(1-g)^2 - 2g(1-g^N)/[N(1+g)(1-g)^3] \\ m = 1/(1-g)^2 \end{array} \right.$$

Notice that the restriction $|g| < 1$ is equivalent to requiring

$$(2.29) \quad B(z) = 1 - gz = 0$$

to have its root outside of the unit circle.

Figure 1 shows the ratio

$$(2.30) \quad q = [\text{Var}(\bar{X}_N)/V_N]^{1/2}$$

for several values of g and varying sample sizes N . The square root comparison is appropriate, for it is the standard deviation of the sample mean that determines the width of a confidence interval for the population mean. Notice that for $|g| \leq 0.50$, the error of approximation is less than 5 percent for $N > 25$. For $|g| = 0.95$, the error is about 10 percent for $N = 100$.

The error patterns differ noticeably for positive and negative values of g . For positive values, m always overestimates $N \text{ Var}(\bar{X}_N)$; for negative values, it always underestimates it. Also, $N \text{ Var}(\bar{X}_N)$ oscillates when $g < 0$, the pattern being most apparent for small N . From inspection of Fig. 1, it is clear that choosing an even value of N improves the approximation.

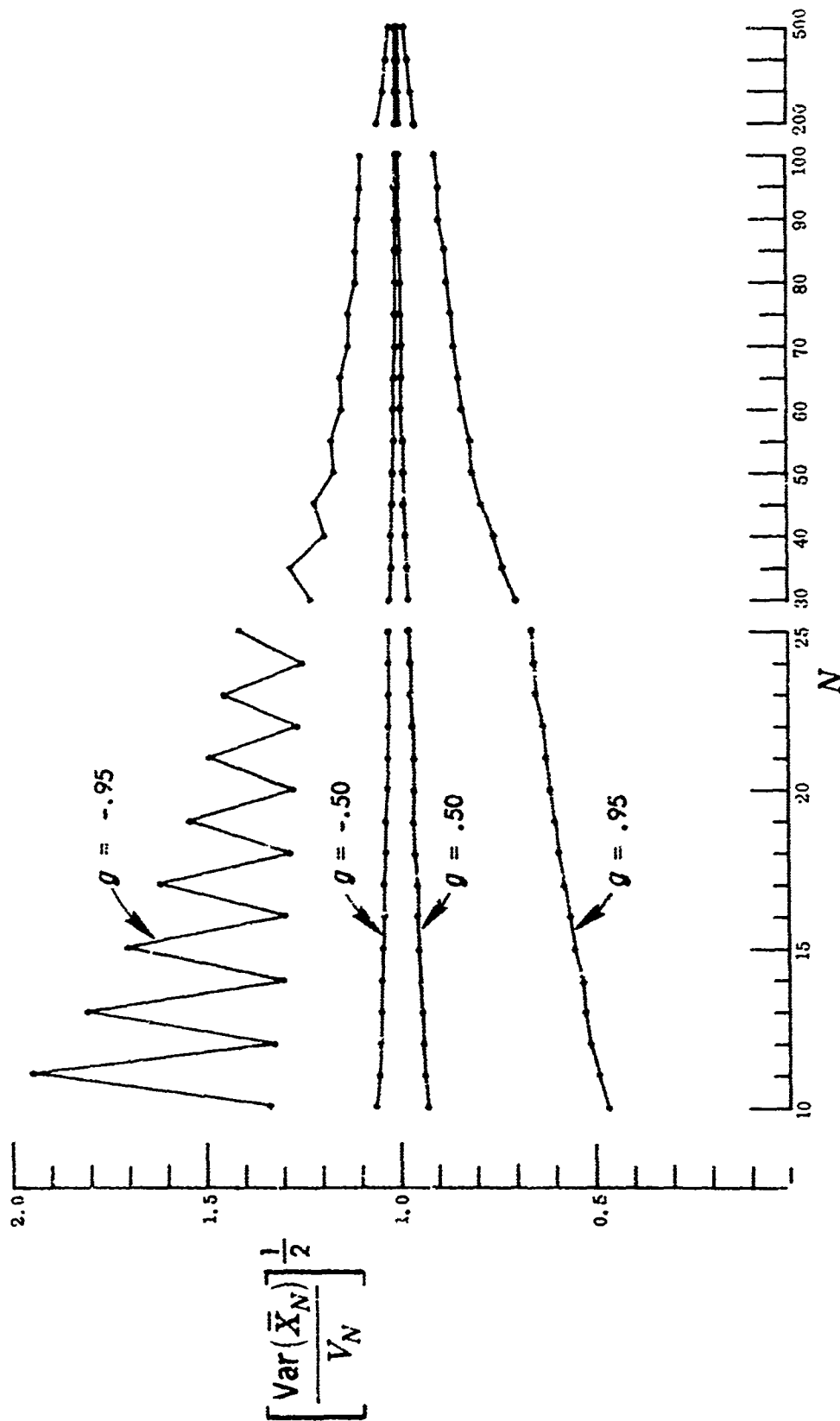


Fig. 1--Comparison of sample mean variance and approximation for a Markov process

Suppose we write

$$Z_t = \sum_{s=0}^r b_s X'_{t-s} = - \sum_{s=r+1}^p b_s X'_{t-s} + Y_t ,$$

(2.31)

$$b_0 = 1; b_s = 0, s > p ,$$

so that

$$(2.32) \quad E(X'_{t-r-1} Z_t) = - \sum_{s=0}^p b_s E(X'_{t-s} X'_{t-r-1}) = - \sum_{s=0}^p b_s R_{s-r-1} .$$

Here Z_t is a conditional random variable derived by removing the linear effects of $X'_{t-1}, \dots, X'_{t-r}$ from X'_t . If $r = p$, then all the linear effects of past events have been removed from X'_t so that

$$E(X'_{t-p-1} Z_t) = 0 ,$$

(2.33)

$$b_{p+1} = 0 .$$

This result is important, for it gives us a way of determining p , the order of the autoregressive representation. Suppose we estimate the coefficients in R autoregressive schemes of orders $1, \dots, R$. We then test the significance of the last estimated coefficient in each scheme. By choosing R sufficiently large, we can find a value $p < R$ beyond which all remaining last estimates are insignificant.

It is important to check a number of schemes for significance rather than proceeding stepwise, successively testing the last estimate, and stopping when the coefficient is insignificant. It may occur that a coefficient b_r for $R < p$ vanishes. Proceeding in a stepwise fashion would cause us to stop testing prematurely.

3. STATISTICAL INFERENCE

Consider a set of N observations, X_1, \dots, X_N , that we wish to fit to an autoregressive representation:

$$(3.1) \quad \sum_{s=0}^r b_s (X_{t-s} - \mu) = Y_t.$$

We first adjust for the sample mean,

$$(3.2) \quad X'_t = X_t - \bar{X}_N,$$

and then compute the sample autocovariances,

$$(3.3) \quad C_{N,\tau} = (1/N) \sum_{t=1}^{N-\tau} X'_t X'_{t+\tau} \quad \tau = 0, 1, \dots, R.$$

The quantity R is the highest order of the autoregressive schemes that we plan to test. Since we plan to compute estimates for schemes of successively higher order, we use a time-conserving, recursive estimation procedure suggested by Durbin [3], which Whittle also describes [15, p. 37].

Let $\hat{b}_{r+1,s}$ be the s^{th} lagged estimated coefficient in the scheme of order $r+1$, and let

$$(3.4) \quad \hat{b}_{r+1,0} = 1 \quad r = 0, 1, \dots, R-1.$$

Then from the r^{th} order scheme, where $r = 0, 1, \dots, R-1$, we compute

$$(3.5) \quad \left\{ \begin{array}{l} v_r = \sum_{s=0}^r \hat{b}_{r,s} c_{N,s} \\ w_r = \sum_{s=0}^r \hat{b}_{r,s} c_{N,r-s+1} \\ \hat{b}_{r+1,r+1} = -w_r / v_r \\ \hat{b}_{r+1,s} = \hat{b}_{r,s} + \hat{b}_{r+1,r+1} \hat{b}_{r,r-s+1} \quad s = 1, \dots, r. \end{array} \right.$$

The sample residual variance that is an estimate of σ^2 is

$$(3.6) \quad \hat{\sigma}_{r+1}^2 = (N - r^{-1})^{-1} \sum_{t=r+1}^N \left[\sum_{s=0}^{r+1} \hat{b}_{r+1,s} (X_{t-s} - \bar{X}_N) \right]^2 \quad r = 0, \dots, R-1.$$

Using a result based on Whittle [15, pp. 72-73], one may show that for large N

$$\begin{aligned} \text{var}(\hat{b}_{r,i}) &\sim w_r / N \\ w_r &= 1 - b_{r,r}^2. \end{aligned}$$

If the confidence interval

$$(3.7) \quad \hat{b}_{r,r} \pm P(\omega_r / N)^{\frac{1}{2}}$$

covers zero, then we accept the hypothesis that $b_{r,r} = 0$. The quantity P is the point on the normal curve corresponding to a significance level α , where

$$(3.8) \quad (2\pi)^{-\frac{1}{2}} \int_{-\infty}^P e^{-x^2/2} dx = 1 - \alpha/2 .$$

This method of determining p is based on the testing procedure described by Jenkins and Watts in [8, pp. 189-200]. There the test statistic has the Student t distribution but, in the present context, we assume N to be sufficiently large so that the normal approximation is acceptable. Quenouille [13] has described an alternative and more precise large sample "goodness of fit" test for autoregressive schemes. Unfortunately, his test requires several more complicated computations than those described above, and it does not appear to be easily incorporated into a simulation program.

Suppose that $b_{r,r}$ is tested for significance for $r = 1, \dots, R$, and that the coefficient $b_{p,p}$, $p < R$ is significant, but the coefficients $b_{r,r}$ are not significant for $r = p + 1, \dots, R$. Then we choose the order of the scheme to be p and estimate m by

$$(3.9) \quad \hat{\sigma}_p^2 / \left(\sum_{s=0}^p \hat{b}_{p,s} \right)^2 .$$

Three questions remain to be answered here. One is the choice of R ; the second, the choice of the initial sample size; and the third, the choice of α and thereby P . The larger is R , the better is the chance of including the correct p within the schemes tested. But low-

order autoregressive schemes often suffice to account for the autocorrelation structure in X and, consequently, choosing R to be 10 should be more than adequate. In our work we have chosen R to be 4 and 5.

The steps that we have so far described are based on an initial sample of N observations. For the normal approximation to be acceptable, we require that $N - R > 30$ so that the initial sample size N exceeds $R + 30$; how much greater it should be depends on the cost of collecting observations, a point to which we return in Sec. 6.

Earlier we spoke of testing the last coefficient in each of the R autoregressive schemes estimated. We then have a multiple testing problem and, if we choose a significance level α for each test, the significance level for the multiple test will be greater than α . The greater R is, the greater the divergence is between α and the significance level for the multiple test. It is therefore recommended that the choice of α be less than one would customarily use in testing a single hypothesis. This divergence is also a good reason for keeping R small.

4. RELIABILITY CRITERIA

The purpose of the present research is to determine the sample size needed to estimate the population mean with a specified accuracy or reliability. Since we are treating the sample mean \bar{X}_N as a normal variate with approximate variance m/N , we specify reliability by means of the confidence statement

$$(4.1) \quad \Pr(|\bar{X}_N - \mu| < Q\sqrt{m/N}) \sim 1 - \beta ,$$

where the confidence level β is a small probability such as 0.05 or 0.10, and Q is the normal point corresponding to

$$(2\pi)^{-1/2} \int_{-\infty}^Q e^{-z^2/2} dz = 1 - \beta/2 .$$

We may also write (4.1) as

$$(4.2) \quad \Pr(\bar{X}_N - Q\sqrt{m/N} < \mu < \bar{X}_N + Q\sqrt{m/N}) \sim 1 - \beta ,$$

and we note that the larger is N , the shorter is the confidence interval around μ .

Suppose we wish to collect a sample size such that the variance of the resulting sample mean is less than or equal to V with probability $1 - \beta$. That is,

$$(4.3) \quad \Pr(\bar{X}_{N^*} - Q\sqrt{V} < \mu < \bar{X}_{N^*} + Q\sqrt{V}) \sim 1 - \beta .$$

To determine N^* , we note the equivalence of (4.2) and (4.3) so that

$$(4.4) \quad V = m/N^*,$$

$$(4.5) \quad N^* = m/V.$$

If m were known *a priori*, then the determination of N^* in (4.5) would follow directly. Since we only have an estimate \hat{m} of m , it is natural to replace m by \hat{m} in (4.5) and so estimate N^* . As a first approximation, this is a reasonable approach. Figure 2 shows a flow chart of an iterative procedure for determining and collecting N^* observations.

Notice that we collect $\gamma(N^* - N)$, not $N^* - N$ additional observations. This scaling increases the number of iterations, but, more important, it decreases the total number of unnecessary observations collected. Since the computer time saving will generally be much greater with regard to the avoidance of collecting unnecessary observations than the time expended on additional iterations, scaling is desirable. In the examples to be presented in Sec. 8, γ was set equal to $1/3$ and $1/2$ for comparison.

The specification of V is a statistically oriented constraint in terms of population parameters. Often we prefer a confidence statement

$$(4.6) \quad \Pr(\bar{X}_{N^*} - c < \mu < \bar{X}_{N^*} + c) \sim 1 - \beta,$$

where c is a specified constant. Here we wish to determine a sample size N^* such that the probability is about $1 - \beta$ that the difference between the sample and population means does not exceed $\pm c$. This is an absolute reliability criterion to be met. To determine N^* , we

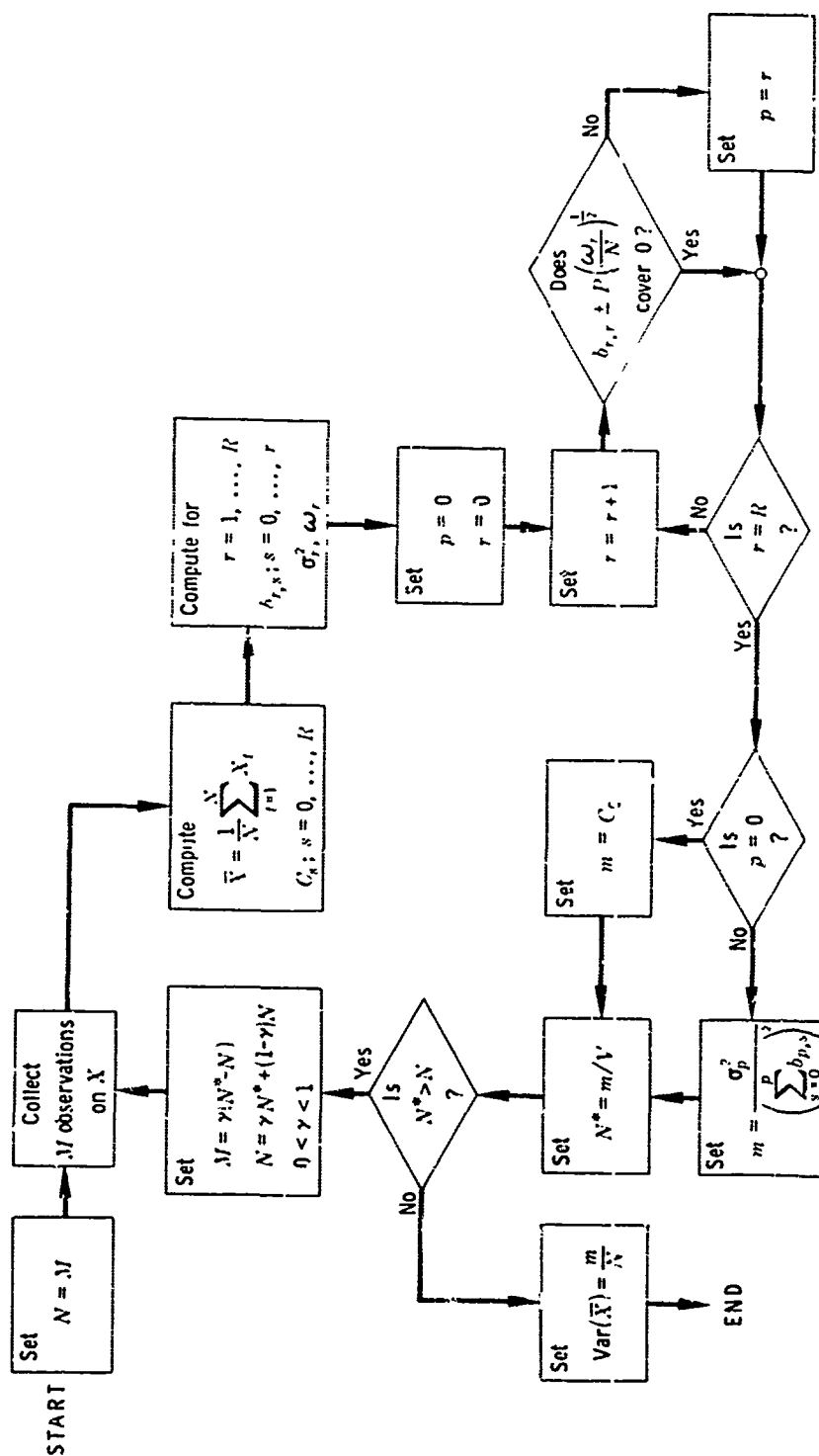


Fig. 2--Computation of unweighted sample mean (M, P, R, V and γ specified)

note that (4.2) and (4.6) are identical for

$$(4.7) \quad c = Q\sqrt{m/N^*} ,$$

so that

$$(4.8) \quad N^* = m(Q/c)^2 .$$

And we note that the relationship between V and c is

$$V = (c/Q)^2 .$$

As an alternative, we may wish to determine N^* such that

$$(4.9) \quad \Pr(\bar{X}_{N^*} - c\mu < \mu < \bar{X}_{N^*} + c\mu) \sim 1 - \beta ,$$

so that the probability is $1 - \beta$ that the difference between the sample and population means does not exceed $\pm c\mu$. Using (4.2) again, we have

$$(4.10) \quad N^* = m(Q/c\mu)^2 .$$

This is a relative reliability criterion. Notice in (4.8) and (4.10) that halving c causes a fourfold increase in N^* . Here m and μ are unknown, so we replace them in (4.10) by \hat{m} and \bar{X}_{N^*} .

Both the absolute and relative criteria are being determined in (4.8) and (4.10) using Q from the normal distribution. In the absolute case we know from theoretical considerations that failure to account for the substitution of \hat{m} for m makes Q smaller than it should be and, therefore, causes an underestimate of N^* . For the relative case, the

problem is further compounded by the substitution of \bar{X}_{N^*} for μ .

By drawing several analogies to the case of independent observations with unknown mean and variance, we can introduce a correction factor for Q to account for the unknown m . This is done in the next section.

5. IMPROVED CONFIDENCE INTERVALS

The mean and variance of the process under study are μ and R_0 , respectively. Let us now consider a hypothetical process made up of independent events each with mean μ and variance R_0 . For L observations the variance of the sample mean of the hypothetical process is R_0/L . In the process under study it is m/N for large N . Equating these sample mean variances, we have

$$(5.1) \quad R_0/L = m/N ,$$

so that, with regard to the sample mean variance,

$$(5.2) \quad K = N/L = m/R_0$$

is the number of observations to be collected on the process under study that is equivalent to collecting one independent observation on the hypothetical process.

Suppose we have an estimate C_0 of R_0 for the hypothetical process. Then we use the t distribution with $L - 1$ degrees of freedom, together with \bar{X}_L and C_0/L , to compute a confidence interval for μ . The use of $L - 1$ instead of L is due to the substitution of \bar{X}_L for μ in C_0 . In the present study we may derive a more representative confidence interval for μ by using the t distribution with \bar{X}_N , and C_0/L with $L - 1$ equivalent degrees of freedom, where

$$(5.3) \quad L = N/K = NR_0/m ,$$

the loss of one degree of freedom being for the sample mean substitution.

To estimate L , we replace m in (5.3) by \hat{m} .

To incorporate a table of the appropriate critical values for the t distribution into a computer program seems undesirable since this would require a value for each number of degrees of freedom for each β . Instead, we may use formulae for the asymptotic expansion of the critical value of t around the critical value of the normal distribution for a given β . The interesting characteristic of the asymptotic expansion is that it is a power series in inverse powers of the number of degrees of freedom so that we may compute Q for a given β simply by inserting the number of degrees of freedom in the formulae. These formulae may be found in [14, p. 948].

It is to be noted that the use of the t distribution corrects for unknown m . At the present stage of research no correction can be offered for the unknown μ in the relative reliability case. Nor is there any adjustment for the use of an estimate for L . To check on the extent of degradation due to these omissions, our examples are based on the relative reliability criterion.

6. BIAS ADJUSTMENT

At the beginning of the simulation experiment, a number of vital variables are assigned predetermined values to "prime" the system. This procedure establishes a set of initial conditions and, because of the inherent dependence among events X_1 , the first observation on the process of interest is a function of these initial conditions. The second observation X_2 is also a function of these values but to a lesser extent than X_1 is. Successive observations are less dependent on the initial conditions so that eventually events in the simulation experiment are independent of them.

Because of their dependence on the initial conditions, observations near the beginning of the experiment are not representative of the process of interest and their inclusion in \bar{X}_N makes this quantity a biased estimator of the true mean μ . As N becomes large, the bias goes to zero since early observations become less influential on the average. But for moderate N , the bias may be significant and should be reduced if possible.

We noted that K in (5.2) essentially measures the number of auto-correlated observations per independent observation. Intuitively, we therefore expect the correlation between observations K units apart to be low. For example, in the first-order process described earlier we have

$$(6.1) \quad K = (1 + g)/(1 - g) ,$$

so that the correlation between the first and $K + 1$ st observations is

$$(6.2) \quad g^{(1+g)/(1-g)} < 1/e^2 = .13534 .$$

As a first step toward reducing the influence of initial conditions, we remove the first K observations from the sample. Then we have the sample mean

$$(6.3) \quad \bar{X}_{N,K} = \frac{1}{N-K} \sum_{t=K+1}^N X_t ,$$

with

$$(6.4) \quad V_{N,K} = m/(N-K) = \sigma^2 / [(N-K)b^2] = \sigma^2 / [Nb^2 - \sigma^2 / R_0] .$$

The confidence interval for μ is then computed using the t distribution with $L - 2$ "equivalent" degrees of freedom and $\bar{X}_{N,K}$ and $C_0/(L-1)$, L being estimated as before.

Figure 3 is a flow chart that illustrates one way of including the bias adjustment. When the sample size is judged sufficient, the sample mean is recomputed using the newly estimated bias adjustment K. The estimates \hat{m} and \hat{L} are not recomputed, since our experience has shown that a recomputation of these quantities makes little difference.

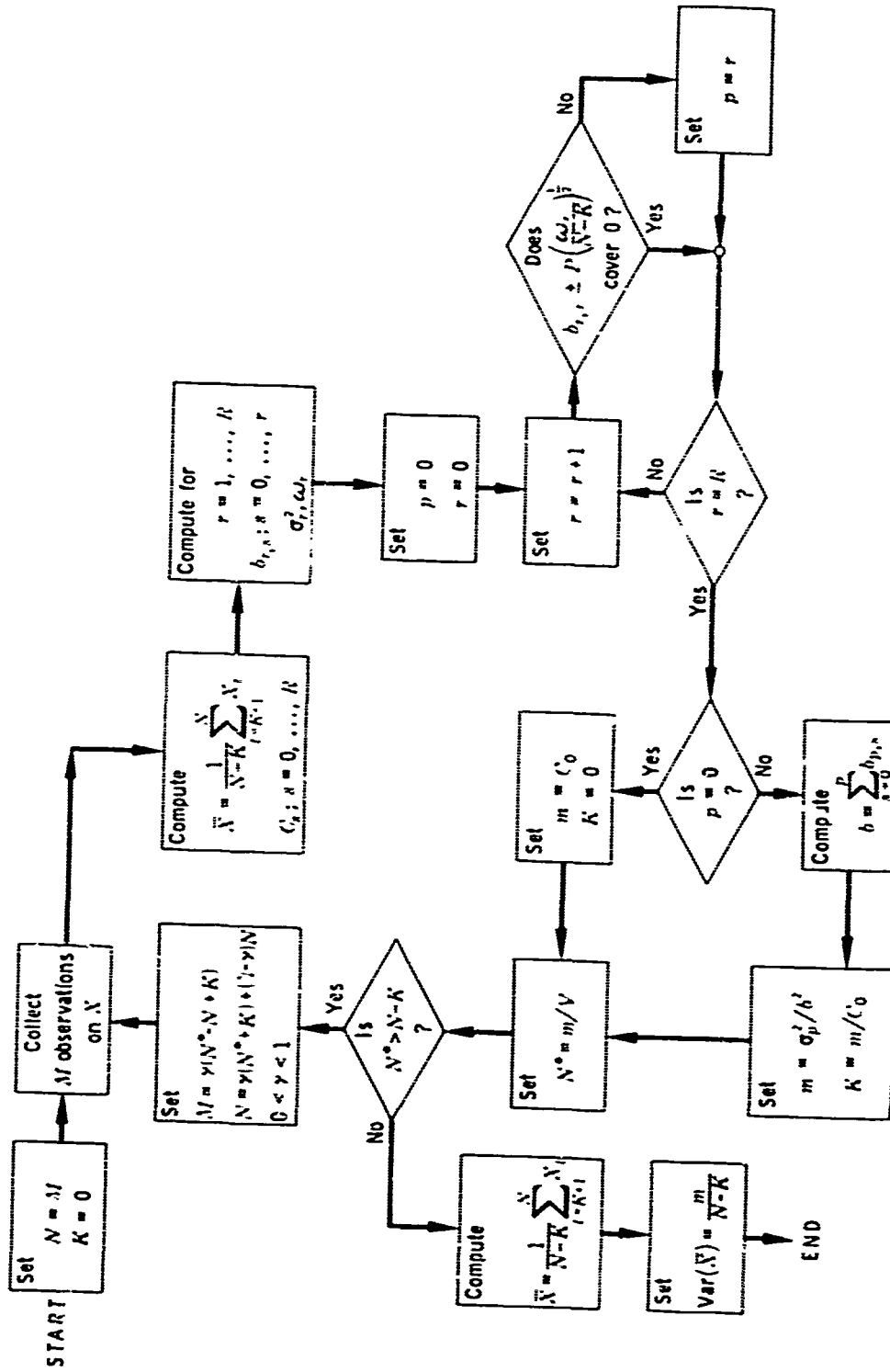


Fig. 3--Computation of unweighted sample mean with bias adjustment (M , P , R , V and γ specified)

7. EXAMPLES

This section presents four examples to illustrate how the proposed technique works. The first three examples are zero-, first- and second-order autoregressive schemes for normal stochastic sequences. One-hundred replications were collected on each example to enable meaningful comparisons for different significance levels α in determining the order p of the autoregressive schemes and for different values of Y , the scale factor used in collecting additional observations. The fourth example is a first-come, first-served single-server queueing problem with independent and exponentially distributed interarrival and service times. The purpose of this example is to study a problem more closely akin to those usually analyzed in discrete event simulation experiments and also one in which the underlying distributions are not normal. Analytical solutions are available for all four examples and serve as a check on the technique. The computer program was written in the SIMSCRIPT II programming language [10].

In the first example we considered a stochastic sequence where

$$(7.1) \quad X_t = 0.5 + Y_t,$$

Y being normal with

$$(7.2) \quad \begin{aligned} E(Y_t) &= 0 \\ E(Y_t Y_s) &= \begin{cases} 1 & t = s \\ 0 & t \neq s \end{cases} \end{aligned}$$

Then

$$(7.3) \quad E(X_t) = \mu = .5 ,$$

and for a sample of N observations,

$$(7.4) \quad \text{Var}(\bar{X}_N) = 1/N .$$

In the second example we considered a first-order scheme

$$(7.5) \quad X_t = 0.5X_{t-1} + .5 + Y_t ,$$

so that

$$(7.6) \quad E(X_t) = \mu = 1 ,$$

$$\text{Var}(\bar{X}_N) \sim 4/N .$$

And for the third example we studied a second-order scheme,

$$X_t = 0.5X_{t-1} - .25X_{t-2} + 0.5 + Y_t ,$$

$$(7.7) \quad E(X_t) = \mu = 2/3 ,$$

$$\text{Var}(\bar{X}_N) \sim 16/(9N) .$$

The quantities $\text{Var}(\bar{X}_N)$ were computed using (2.26). The coefficients in the second-order scheme were chosen to illustrate how a higher-order autoregressive representation does not necessarily imply more autocorrelation in the sequence and, hence, a larger $\text{Var}(\bar{X}_N)$.

The objective in all three examples was to obtain sample sizes N^* such that

$$\Pr(|\bar{X}_{N^*} - \mu| < c\mu) \sim 1 - \beta .$$

$$(7.8) \quad c = 0.20 ,$$

$$\beta = 0.10 .$$

Using (4.5), the required sample size for each example was

$$(7.9) \quad N^* = (16.4)^2 \sim 269 ,$$

so that the respective sample mean variances were

$$\text{Var}(\bar{X}_{N^*}) = 0.003717 ,$$

$$(7.10) \quad \text{Var}(\bar{X}_{N^*}) \sim 0.014870 ,$$

$$\text{Var}(\bar{X}_{N^*}) \sim 0.006609 .$$

For each example, two significance levels for determining the autoregressive order were studied. They were $\alpha = 0.025$ and $\alpha = 0.05$. Also, two scaling factors were examined for determining the number of additional observations to be collected. They are $\gamma = 0.5$ and $\gamma = 0.3333$. Therefore each example contained four cases.

The results for the total of twelve cases, each containing 100 independent replications, are presented in Table 1, where

$p \equiv$ order of the autoregressive scheme (Sec. 2)

$\alpha \equiv$ significance point in test to determine p (Sec. 3)

$\gamma \equiv$ weighting factor for newly computed required sample size (Sec. 5)

$V_N \equiv$ approximation to variance of sample mean (Sec. 2) .

In all experiments, V_N was close to $\text{Var}(\bar{X}_N)$.

Column 4 lists the average required sample size computed on the last iteration of each experiment. Column 5 lists the average sample size collected on the last iteration, which is naturally greater than the corresponding quantity in col. 4 since it is precisely this condition that terminates the experiment. The quantities in parentheses are the sample standard deviations. The highest order autoregressive scheme R considered was 4.

Notice that increasing α from 0.025 to 0.05 causes slightly less than a doubling in col. 8 for $p = 0$. A less marked increase occurs for $p = 1, 2$. Also noteworthy is the general increase in col. 8 for a given α as p increases. These increases would be, in fact, larger if R were greater since more tests would be performed. Since we expect $p > 0$, it is advisable to make α small and also to restrict R . The choices of $\alpha = 0.025$ and $R \leq 4$ appear to be acceptable operating conditions in the cases described.

When γ is reduced from 0.5 to 0.333, the excess sample size (col. 6) becomes smaller. Compared with the theoretical N^* of 269, the use of $\alpha = 0.5$ causes an average of 1.1746 observations to be collected for every required observation, whereas the use of $\alpha = 0.333$ requires an

Table 1
TEST RESULTS FOR 100 REPLICATIONS
(Theoretical $N = 269$)

(1) p	(2) α	(3) v	Average				(8) Probability of Choosing p Incorrectly	(9) Sample V_N	(10) Theoretical V_N
			(4) Final N^* (Required)	(5) Final N (Collected)	(6) Final $N-N^*$	(7) No. of Iterations			
0	0.025	0.5000	277 (85.0)	323 (148.3)	46	4.21	0.0546	0.003448	0.003718
		0.3333	269 (87.9)	275 (96.8)	6	6.27	0.0622	0.003771	
	0.05	0.5000	268 (92.2)	313 (181.6)	45	4.25	0.1059	0.003593	
		0.3333	269 (78.7)	283 (100.5)	14	6.48	0.1755	0.003604	
1	0.025	0.5000	253 (107.1)	301 (163.7)	42	4.12	0.1553	0.013965	0.014872
		0.3333	245 103.4	263 (132.8)	18	5.66	0.1555	0.014843	
	0.05	0.5000	254 (108.7)	310 (195.2)	56	4.01	0.1845	0.013941	
		0.3333	247 (104.7)	260 (127.2)	13	5.39	0.1855	0.014904	
2	0.025	0.5000	265 (94.4)	335 (197.2)	70	3.98	0.2136	0.005981	0.006129
		0.3333	264 (100.5)	287 (136.6)	23	5.62	0.1904	0.006505	
	0.05	0.5000	260 (97.4)	312 (177.5)	53	3.63	0.2342	0.006169	
		0.3333	264 (87.2)	277 (112.0)	13	5.50	0.1636	0.006546	

average of 1.0195 observations. The sample standard deviations in col. 5 are also smaller for $\gamma = 0.3333$, but the average number of iterations (col. 7) is notably increased. Moreover, 27 percent more CPU time is required for $\gamma = 0.3333$. These facts suggest that smaller γ 's improve statistical precision but require more CPU time. We now study the cost of this improved precision.

The actual generation of data used in these experiments consumed relatively little CPU time so that we may reasonably attribute the total CPU time to the proposed statistical technique. For $\gamma = 0.5$, the program processed 274 observations per second; for $\gamma = 0.3333$, it processed 187 observations per second. In all experiments summarized in Table 1, the data were retained in the magnetic core storage unit.

Suppose we theoretically require a sample of 10,000 observations. For $\gamma = 0.5$ this would result in the collection of 11,746 observations, and for $\gamma = 0.3333$, 10,195 observations would be collected. Dividing by the processing times per observation, we note that 42.9 and 54.2 seconds are consumed when $\gamma = 0.5$ and $\gamma = 0.3333$, respectively. Compared to the time consumed in most simulation experiments, this difference for the two values of γ is negligible.

The fourth example is a single-server queueing problem with independent and exponentially distributed interarrival and service times and a first-come, first-served queueing discipline. The mean interarrival and service times are $\lambda_1 = 0.25$ and $\lambda_2 = 0.225$, respectively, so that the activity level is

$$(7.11) \quad \rho = \lambda_2 / \lambda_1 = 0.9 .$$

From theory we know that the mean queue length is

$$(7.12) \quad \mu = \rho / (1 - \rho) = 9 ,$$

and its variance is

$$(7.13) \quad R_0 = \rho / (1 - \rho)^2 = 90 .$$

Using the results of [5 and 12], the variance of the sample mean derived from observing queue length for a time interval $(0, T)$ is for large T

$$(7.14) \quad \text{Var} (\bar{X}_T) \sim \rho(1 + \rho)\lambda_2 / [(1 - \rho)^4 T] = 3847.5/T ,$$

where T is measured in the same time units as λ_1 and λ_2 .

We recorded queue length at $N = T$ unit intervals. Since some variation is eliminated by this discrete sampling, we may regard $3847.5/N$ as an upper bound on the variance of \bar{X}_N , the resulting sample mean.

As a reliability criterion we chose

$$(7.15) \quad \Pr (|\bar{X}_{N^*} - \mu| < c\mu) = 1 - \beta ,$$

where

$$(7.16) \quad \begin{aligned} c &= .3 , \\ \beta &= .10 . \end{aligned}$$

Then using (4.5), the upper bound on N^* was 1420. Three sets of replications were run, each with

$$(7.17) \quad \begin{aligned} \alpha &= 0.025, \\ \lambda &= 0.3333, \\ R &= 5. \end{aligned}$$

R was chosen to emphasize that it need not be fixed at 4 as in the previous examples. The initial sample size in the first set was $M = 250$, in the second $M = 500$, and in the third $M = 1000$.

The relevant sample averages are presented in Table 2. Two tendencies are noteworthy. The first is the constant underestimation of the confidence interval width. This is no doubt partially due to the use of sampling theory appropriate for an absolute reliability criterion with a relative criterion. The second point is the improved estimate of sample size as the initial sample size gets larger. Clearly the choice of initial sample size makes a difference in the final result.

Table 2
AVERAGES FOR QUEUEING PROBLEM

	Theoretical	Sample		
		M=250	M=500	M=1000
Mean	9	8.21	8.15	8.83
Lower confidence point	6.3	5.86	5.89	6.41
Upper confidence point	11.7	10.56	10.42	11.22
Interval width	5.4	4.70	4.53	4.81
Variance of sample mean	2.71	2.10	1.90	2.37
Final sample size	1462 ^a	1068	1299	1473
Equivalent degrees of freedom	32	28	33	30
Bias adjustment	43	32	37	46

^aIncludes bias adjustment.

To measure the overall adequacy of the suggested procedure, we studied the number of times that he generated confidence intervals included the true mean of 9. These results are presented in Table 3, where they are dichotomized by replications that terminated after at most one iteration and by those that did not. That is, if the experiment terminated on the first statistical analysis, then no iterations occurred.

Table 3

CONFIDENCE INTERVALS FOR QUEUEING PROBLEM

M	Iterations	Replications	Include Mean		Sample Divided by Expected
			Sample	Expected	
250	≤ 1	29	8	26.1	0.31
	> 1	71	58	63.9	0.81
	Total	100	66	90.0	0.73
500	≤ 1	36	15	32.4	0.46
	> 1	64	56	57.6	0.97
	Total	100	71	90.0	0.79
1,000	≤ 1	54	36	48.6	0.74
	> 1	46	43	41.4	1.04
	Total	100	79	90.0	0.88
250	Modified	28	18	25.2	0.71
	Normal	72	57	64.8	0.88
	Total	100	75	90.0	0.83

Since $\beta = 0.1$, we expect the confidence interval to include the mean in 90 percent of the replications. The last column shows the ratio of actual to expected inclusions. Notice that for zero or one iteration the results are considerably poorer than for greater than one iteration. As the initial sample size increases, there is clear improvement in the results for zero or one iteration as well as for the remaining category.

A priori we generally have little knowledge of what an appropriate sample size is for the experiment to be undertaken. Therefore we should be suspicious of results generated by zero or one iteration since early termination implies that our initial guess is close to the correct answer, a remote possibility.

One way to guard against premature stopping is to require a minimum of two iterations per experiment. To test the effect of this constraint, we performed the one-hundred replications again with $M = 250$, with the added requirement that if any experiment terminated with $N - K$ useful observations on the zeroth or first iteration we added $N - K$ to the required sample size and continued the experiment. The results are shown in the last three rows of Table 3 and are to be compared with the corresponding first three rows. Note the significant improvement in experiments that formerly required less than two iterations. It is clear that an *ad hoc* rule such as the one just described improves the statistical properties of the experimental results.

As Table 3 shows, the percentage of confidence intervals that include the mean generally falls below that expected from theory. Apart from premature stopping, we attribute this difference partially to the use of a relative criterion with a distribution theory for absolute criterion, and to the approximation involved in using theory developed for independent observations for autocorrelated ones.

One remaining and no doubt crucial consideration is the assumption of normality of \bar{X}_{N^*} . We have used it because it is intuitively plausible and convenient. It is therefore instructive to study the sample size problem without this assumption to understand the advantages of the normality assumption. Using Chebyshev's Inequality, we have in the present setting

$$(7.18) \quad \Pr (|\bar{X}_N - \mu| \leq k\sqrt{m/n}) \geq 1 - 1/k^2 .$$

For the queueing problem we have

$$cu = k\sqrt{m/n} = 2.7 ,$$

$$(7.19) \quad m = 3847.5 ,$$

$$N^* = 1420 ,$$

so that

$$1/k^2 = \frac{m}{(2.7)^2 N^*} = 0.3704 ,$$

(7.20)

$$\Pr (|\bar{X}_{N^*} - \mu| \leq 2.7) \geq 0.6296 .$$

This means that, regardless of the distribution of \bar{X}_{N^*} , the probability is at least 0.6296 that deviations from the mean are no greater than ± 2.7 . Our results in Table 3 are clearly better than this low boundary.

Suppose we wish to make no distributional assumption about \bar{X}_N and to use Chebyshev's Inequality to determine the required sample size N^* for $1 - 1/k^2 = 0.90$. Then

$$k^2 = 10 ,$$

so that

$$\Pr (|\bar{X}_{N^*} - \mu| \leq 2.7) \geq 0.9 ,$$

$$(7.21) \quad k^2_{m/N^*} = (2.7)^2 ,$$

$$N^* = \frac{k^2_m}{(2.7)^2} = \frac{(10)(3847.5)}{(2.7)^2} = 5278 .$$

Therefore, we require an additional $(5278-1420) = 3858$ observations to meet the probability requirement. In general, the use of Chebyshev's Inequality is a luxury that most users prefer to forego, given the intuitive plausibility of the normality assumption.

8. THE MINIMUM VARIANCE ESTIMATOR

In Sec. 6, a knowledge of the autoregressive coefficients $\{b_s; s = 0, \dots, p\}$ enabled us to reduce the bias in the sample mean \bar{X}_N . It is only natural to inquire whether or not a knowledge of these coefficients can enable us to derive an estimator of the population mean μ that has smaller variance than \bar{X}_N has for a given sample size N . The answer is yes and it is, in fact, possible to derive the minimum variance estimator, which we do here. To develop the idea, we neglect the bias due to initial conditions and correct for it later.

Consider the mean estimator

$$(8.1) \quad \tilde{X}_N = \sum_{t=1}^N \theta_t X_t,$$

such that

$$(8.2) \quad \sum_{t=1}^N \theta_t = 1$$

Then

$$(8.3) \quad E(\tilde{X}_N) = \mu,$$

$$\text{Var}(\tilde{X}_N) = \sum_{s,t=1}^N \theta_s \theta_t R_{s-t}.$$

We define the $1 \times N$ vector of weights,

$$(8.4a) \quad \underline{\theta}_N = (\theta_1, \dots, \theta_N) ,$$

and the $N \times N$ autocovariance matrix,

$$(8.4b) \quad \underline{\Sigma}_N = \begin{bmatrix} R_0 & R_1 & . & . & . & R_{N-1} \\ R_1 & R_0 & & & & . \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ R_{N-1} & . & . & . & . & R_0 \end{bmatrix} ,$$

so that

$$(8.4c) \quad \text{Var}(\tilde{X}_N) = \underline{\theta}_N \underline{\Sigma}_N^{-1} \underline{\theta}_N' .$$

An array of the form (8.4b) is called a Toeplitz matrix and has a number of desirable properties, one of which will benefit us shortly.

We desire to choose $\underline{\theta}_N$ to minimize the variance (8.4c) subject to the condition (8.2). The weights are then

$$(8.5) \quad \underline{\theta}_N = \frac{\underline{\lambda}_N \underline{\Sigma}_N^{-1}}{\left(\underline{\lambda}_N \underline{\Sigma}_N^{-1} \underline{\lambda}_N' \right)} ,$$

where $\underline{\lambda}_N$ is a $1 \times N$ vector of ones. Then

$$(8.6) \quad \text{Var}(\tilde{X}_N) = 1 / \left(\underline{\lambda}_N \underline{\Sigma}_N^{-1} \underline{\lambda}_N' \right) .$$

The weights are functions of the inverse of $\underline{\Sigma}_N$ and, at first glance, one might expect this array to be beyond us unless we know all N autocovariances in (8.4b). Luckily, a knowledge of the autoregressive coefficients $\{b_s; s = 0, \dots, p\}$ and the residual variance σ^2 suffices to derive $\underline{\Sigma}_N^{-1}$. Let σ^{jk} be the element of the inverse in row $j + 1$ and column $k + 1$. Using a result of Whittle [12, p.73], one may show that

$$\sigma^{jk} = (i/\sigma^2) = \sum_{r=0}^{\infty} (b_{j-r} b_{k-r} - b_{r+N-j} b_{r+N-k}) \quad j, k = 0, \dots, N-1, \quad (8.7)$$

$$b_r = 0 \quad r < 0, r > p.$$

Then the weights are

$$\theta_{k+1} = \theta_{N-k} = \sum_{s=0}^k b_s / \left[\sum_{s=0}^p (N - 2s) b_s \right] \quad k = 0, \dots, p-1, \quad (8.8)$$

$$\theta_{k+1} = b / \left[\sum_{s=0}^p (N - 2s) b_s \right] \quad k = p, \dots, N-p,$$

assuming that $N \geq 2p + 1$. The variance is then

$$\text{Var}(\tilde{X}_N) = \sigma^2 / \left[b \sum_{s=0}^p (N - 2s) b_s \right]. \quad (8.9)$$

At least three points deserve mention. First we note that

$$(8.10) \quad \lim_{N \rightarrow \infty} N \text{ var}(\bar{X}_N) = \sigma^2/b^2,$$

so that \bar{X}_N , the conventional estimator, is asymptotically minimum variance. More important, however, is the second observation that (8.9) is an exact formula for the variance in contrast to using the limit $\sigma^2/(Nb^2)$ for \bar{X}_N . The third point is that for a scheme of order p , observations $p, \dots, N-p$ receive equal weight so that the effect of autocorrelation is felt at the ends of the time series.

To compare \bar{X}_N and \tilde{X}_N , a further example is helpful. Consider the first-order autoregressive scheme (2.30) again. We have

$$(8.11) \quad \begin{aligned} \theta_1 = \theta_N &= 1/[N(1-g) + 2g], \\ \theta_k &= (1-g)/[N(1-g) + 2g] \quad k = 2, \dots, N-1, \end{aligned}$$

$$\text{Var}(\tilde{X}_N) = 1/[N(1-g)^2 + 2g(1-g)].$$

Notice that for $0 < g < 1$ the first and last observations receive the largest weights. Since the first observation is subject to the greatest bias from the initial conditions, it is important to remove this bias when the minimum variance estimator is used. By contrast, we note that for $-1 < g < 0$, the first and last observations receive less weight than the remaining observations.

Let N be the required sample size for a specified reliability using (2.5a), and let \tilde{N} be the required sample size using (8.1)

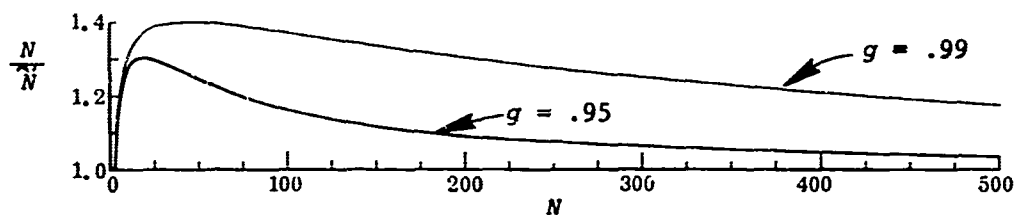


Fig. 4--Comparison of \bar{X}_N and \tilde{X}_N for first-order autoregressive schemes

Figure 4 shows N/\tilde{N} for values of N , and $g = 0.95, 0.99$. For $g = 0.95$, the advantage of \tilde{X}_N is most outstanding for sample sizes of less than 100. The desirability of \tilde{X}_N therefore is most apparent when high unit collection costs require a limit on the sample size. Perhaps more important is the fact that the expression for the variance is known exactly for any order scheme.

The bias adjustment can be introduced here simply by discarding the first K observations and using (8.1), with N replaced everywhere by $N - K$, provided that $N - K \geq 2p + 1$. The new mean estimator is then

$$(8.12) \quad \tilde{X}_{N,K} = \sum_{t=1}^{N-K} \theta_t X_{t+K},$$

with variance

$$(8.13) \quad \text{Var}(\tilde{X}_{N,K}) = \sigma^2 / [b \sum_{s=0}^p (N - K - 2s) b_s].$$

Suppose that it is desired to compute N^* satisfying the absolute reliability criterion (4.1). Using (7.9) we have

$$(8.14) \quad N^* = m(Q/c)^2 + (2/b) \sum_{s=0}^p sb_s .$$

If the relative criterion (4.9) is used, we have

$$(8.15) \quad N^* = m[Q/(c\mu)]^2 + (2/b) \sum_{s=0}^p sb_s .$$

If the only specification is that the variances be less than or equal to V , then

$$(8.16) \quad N^* = m/V + (2/b) \sum_{s=0}^p sb_s .$$

This is the criterion shown in the flow chart in Fig. 5 with the bias correction. It will be helpful to note, when computing the θ_k 's, that

$$(8.17) \quad \begin{aligned} \theta_{K+1} &= \theta_N = \theta' , \\ \theta_{K+1+k} &= \theta_{N-k} = \theta_{K+k} + \theta' b_k & k = 1, \dots, p-1 , \\ \theta_{K+1+k} &= \theta_{K+p} & k = p, \dots, N-p , \end{aligned}$$

$$\theta' = 1 / \left[\sum_{s=0}^p (N - 2s) b_s \right] .$$

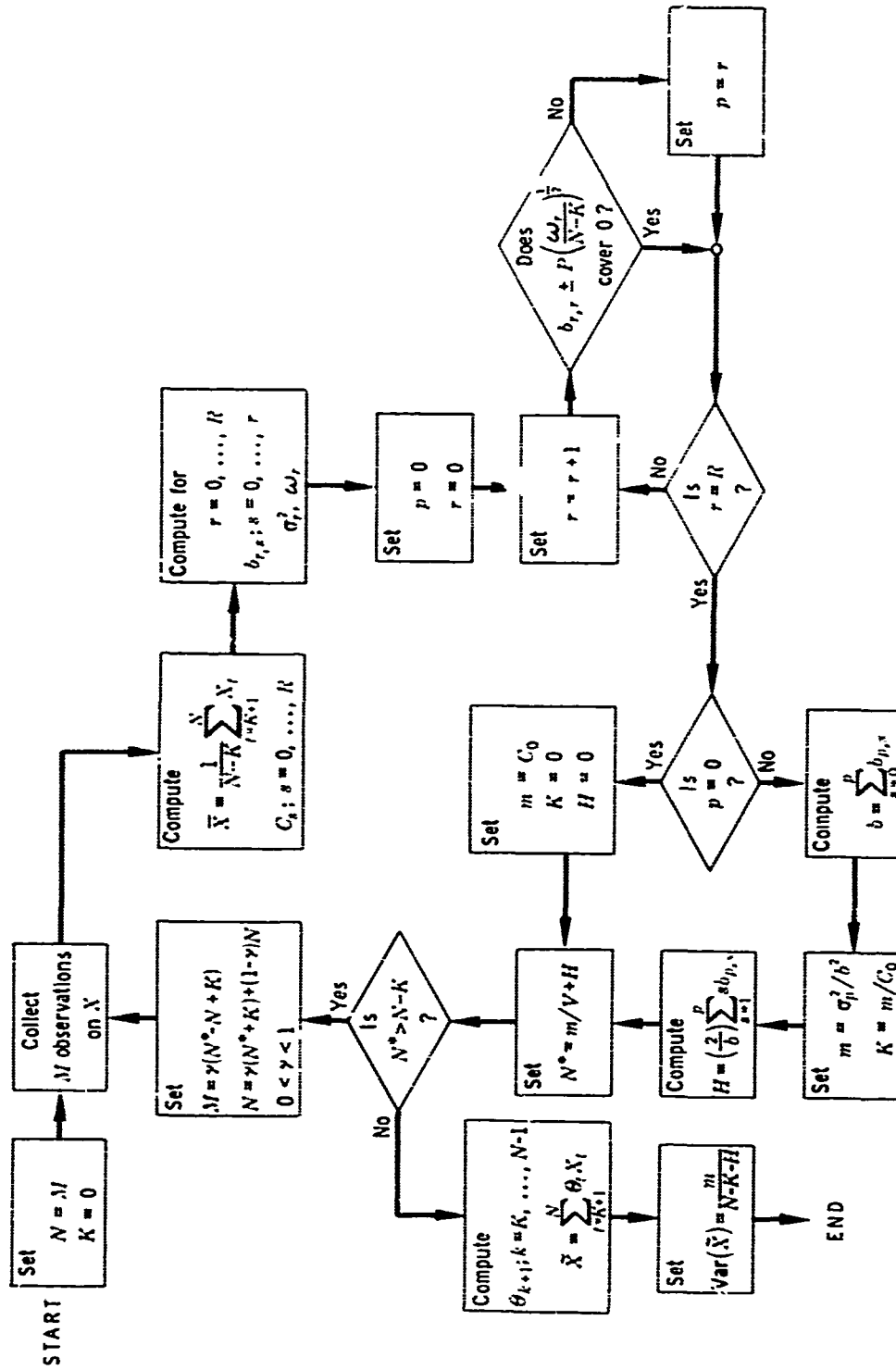


Fig. 5--Computation of minimum variance sample mean with bias adjustment (M, P, R, V and γ specified)

9. CONCLUSIONS

This Memorandum has presented a number of ideas which, when taken together, enable the user to derive results from his simulation experiment with a specified accuracy. Moreover, the results are obtainable without user interactions with the ongoing experiment.

In contrast to the spectral approach [5,6], which often requires the computation of a large number of autocovariances to estimate the variance of the sample mean, the suggested autoregressive approach requires R sample autocovariances where R need not exceed 4 or 5. The autoregressive approach also supplies more objective estimates of the equivalent degrees of freedom that in turn permit the computation of confidence intervals explicitly acknowledging the estimated variance of the same mean. The ability to correct for bias due to initial conditions is also noteworthy.

The Memorandum has dealt with the problem of determining size for estimating the mean. An ancillary problem that deserves attention is how to estimate the mean *efficiently*. To this end, Sec. 8 has described the minimum variance unbiased estimator, which turns out to be a function of the autoregressive coefficients. No doubt such results will interest those concerned with variance reduction techniques.

The flow charts provide procedural assistance for implementing the technique described. The omission of an example using the minimum variance estimator is deliberate, since we prefer to learn more about the resulting sampling properties before using it extensively.

REFERENCES

1. Cox, D. R., and P.A.W. Lewis, The Statistical Analysis of Series of Events, Methuen, London, 1966.
2. Cox, D. R., and H. D. Miller, The Theory of Stochastic Processes, John Wiley and Sons, New York, 1965.
3. Durbin, J., "The Fitting of Time-Series Models," Rev. Int. Inst. Stat., Vol. 28, 1960, pp. 233-244.
4. Fishman, G. S., Problems in the Statistical Analysis of Simulation Experiments: The Comparison of Means and the Length of Sample Records, The RAND Corporation, RM-4880-PR, February 1966. Also published in Comm. ACM, Vol. 10, February 1967, pp. 94-99.
5. _____, and P. J. Kiviat, "The Analysis of Simulation Generated Time Series," Mgt. Sci., Vol. 13, No. 7, March 1967, pp. 525-557.
6. Fishman, G. S., "The Allocation of Computer Time in Comparing Simulation Experiments," Opns. Res., Vol. 16, No. 2, March-April 1968, pp. 280-295.
7. Fraser, D., Nonparametric Methods in Statistics, John Wiley and Sons, New York, 1957.
8. Jenkins, G. M., and D. G. Watts, Spectral Analysis and Its Applications, Holden Day, San Francisco, 1968.
9. Kendall, M. G., and A. Stuart, The Advanced Theory of Statistics, Vol. 3, Hafner, London, 1966.
10. Kiviat, P. J., R. Villanueva and H. M. Markowitz, The SIMSCRIPT II Programming Language, The RAND Corporation, R-460-PR, October 1968.
11. Malinvaud, E., Statistical Methods in Econometrics, Rand McNally, Chicago, 1966.
12. Morse, P., "Stochastic Properties of Waiting Lines," Opns. Res., Vol. 3, No. 3, August 1955, pp. 255-261.
13. Quenouille, M. H., "A Large-Sample Test for the Goodness of Fit of Autoregressive Schemes," J. Roy. Stat. Soc., Series A, Vol. 110, 1947, pp. 123-219.
14. U.S. Department of Commerce, Handbook of Mathematical Functions, National Bureau of Standards, Washington, 1964.
15. Whittle, P., Prediction and Regulation by Linear Least-Square Methods, The English Universities Press, London, 1963.
16. Wold, H., A Study in the Analysis of Stationary Time Series, 2d ed., Almqvist and Wicksell, Stockholm, 1954.

DOCUMENT CONTROL DATA

1 ORIGINATING ACTIVITY THE RAND CORPORATION		2a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	
		2b. GROUP	
3 REPORT TITLE DIGITAL COMPUTER SIMULATION: ESTIMATING SAMPLE SIZE			
AUTHOR(S) (Last name, first name, initial) Fishman, George S.			
REPORT DATE August 1969		6a. TOTAL No. OF PAGES 61	6b. No. OF REFS. 16
CONTRACT OR GRANT No. F44620-67-C-0045		8. ORIGINATOR'S REPORT No RM-5966-PR	
c AVAILABILITY/LIMITATION NOTICES DDC-1		9b. SPONSORING AGENCY United States Air Force Project RAND	
10. ABSTRACT An algorithm for automatically estimating and collecting the sample size required for statistical precision in a computer simulation experiment while the simulation is running. The algorithm, which would be incorporated directly into the computer routines, would relieve an investigator of the burden of first estimating the variance of the sample mean from a data sample obtained from a trial run, then estimating the sample size necessary for the specified confidence interval, and finally collecting that many more observations in a successive simulation run. The underlying probability model is autoregressive: it would depend on an autoregressive representation of the sequence that considers each observation as a linear combination of past observations plus an uncorrelated random residual. This approach need not require more than 4 or 5 autocovariance computations to estimate the variance of the sample mean. A flowchart is included to aid in building the technique into simulation programs.		11. KEY WORDS Computer simulation Statistical methods and processes	